RAIRO-Theor. Inf. Appl. 43 (2009) 767–774
DOI: 10.1051/ita/2009019

Available online at:
www.rairo-ita.org

A NOTE ON THE NUMBER OF SQUARES IN A PARTIAL WORD WITH ONE HOLE*

Francine Blanchet-Sadri[1] and Robert Mercaş[2]

Abstract. A well known result of Fraenkel and Simpson states that the number of distinct squares in a word of length n is bounded by $2n$ since at each position there are at most two distinct squares whose last occurrence starts. In this paper, we investigate squares in partial words with one hole, or sequences over a finite alphabet that have a "do not know" symbol or "hole". A square in a partial word over a given alphabet has the form uv where u is *compatible* with v, and consequently, such square is compatible with a number of words over the alphabet that are squares. Recently, it was shown that for partial words with one hole, there may be more than two squares that have their last occurrence starting at the same position. Here, we prove that if such is the case, then the length of the shortest square is at most half the length of the third shortest square. As a result, we show that the number of distinct squares compatible with factors of a partial word with one hole of length n is bounded by $\frac{7n}{2}$.

Mathematics Subject Classification. 68R15, 05A05.

1. Introduction

A well known problem is the determination of the maximum number of distinct squares in any word of length n. With this problem progress has been made: Fraenkel and Simpson showed that this number is at most $2n$ [6], a result recently proved somewhat more simply by Ilie [7], then later improved to

Keywords and phrases. Combinatorics on words, partial words, squares.

* *This material is based upon work supported by the National Science Foundation under Grant No. DMS-0754154. The Department of Defense is also gratefully acknowledged. We thank Abraham Rashin and Kristen Wetzler for useful suggestions. An Internet server interface at* www.uncg.edu/cmp/research/squares *has been established for automated use of the program.*

[1] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402–6170, USA; blanchet@uncg.edu

[2] GRLMC, Universitat Rovira i Virgili, Departament de Filologies Romàniques, Av. Catalunya 35, Tarragona, 43002, Spain.

Article published by EDP Sciences

$2n - \Theta(\log n)$ [8]. Experiment strongly suggests that this number is less than n. In order to show this, one needs to somehow limit to less than one the average number of squares whose last occurrence begins at the positions of the word.

In this paper, we investigate the problem of counting distinct squares in partial words with one hole, or sequences over a finite alphabet that may contain a "do not know" symbol or "hole". The counting of distinct squares in partial words was recently initiated by Blanchet-Sadri et al. and revealed surprising results [4]. There, it was shown that for partial words with one hole, there may be more than two squares that have their last occurrence starting at the same position, and that if such is the case, then the hole is in the shortest square. As discussed in [4], although computations show that the actual bound for one-hole partial words of length n is at most n distinct squares, the results obtained there using the approach of Fraenkel and Simpson make the bound directly dependable on the size of the alphabet. Finding a dependency between the maximum number of squares starting at one position and the length of the word might be a solution. Solving this problem, at least partially, could also give a new perspective to the study of the maximum number of distinct squares in words without holes.

In Section 2, we will prove that if three squares have their last occurrence starting at the same position in a partial word with one hole, then the length of the shortest square is at most half the length of the third shortest square. As a result, we will show that the number of distinct squares compatible with factors of a partial word with one hole of length n is bounded by $\frac{7n}{2}$. In Section 3, we will conclude with some remarks on the positions where three or more squares have their last occurrences.

In the rest of this section, we review basic concepts on partial words. Fixing a nonempty finite set of letters or an *alphabet* A, a *partial word* u of length $|u| = n$ over A is a partial function $u : \{0, \ldots, n-1\} \to A$. For $0 \leq i < n$, if $u(i)$ is defined, then i belongs to the *domain* of u, denoted by $i \in D(u)$, otherwise i belongs to the *set of holes* of u, denoted by $i \in H(u)$. The unique word of length 0, denoted by ε, is called the *empty* word. For convenience, we will refer to a partial word over A as a word over the enlarged alphabet $A_\diamond = A \cup \{\diamond\}$, where $\diamond \notin A$ represents a hole. The set of all words (respectively, partial words) over A of finite length is denoted by A^* (respectively, $A_\diamond^*$).

The partial word u is *contained in* the partial word v, denoted by $u \subset v$, provided that $|u| = |v|$, all elements in $D(u)$ are in $D(v)$, and for all $i \in D(u)$ we have that $u(i) = v(i)$. The partial words u and v are *compatible*, denoted by $u \uparrow v$, provided that there exists a partial word w such that $u \subset w$ and $v \subset w$. An equivalent formulation of compatibility is that $|u| = |v|$ and for all $i \in D(u) \cap D(v)$ we have that $u(i) = v(i)$. The following rules are useful for computing with partial words: (1) *Multiplication:* if $u \uparrow v$ and $x \uparrow y$, then $ux \uparrow vy$; (2) *Simplification:* if $ux \uparrow vy$ and $|u| = |v|$, then $u \uparrow v$ and $x \uparrow y$; and (3) *Weakening:* if $u \uparrow v$ and $w \subset u$, then $w \uparrow v$. If u, v are nonempty compatible partial words, then uv is called a *square*. Whenever we refer to a square uv it will imply that $u \uparrow v$.

A *period* of a partial word u is a positive integer p such that $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j \bmod p$. In this case, we call u *p-periodic*. A *weak period* of u is a positive integer p such that $u(i) = u(i+p)$ whenever $i, i+p \in D(u)$. In this case, we call u *weakly p-periodic*. Note that every weakly p-periodic word is p-periodic but this is not necessarily true for partial words.

For partial words u, v, w, if $w = uv$, then u is a *prefix* of w, denoted by $u \leq w$, and if $v \neq \varepsilon$, then u is a *proper prefix* of w, denoted by $u < w$. If $w = xuy$, then u is a *factor* of w.

2. Bound on the number of squares

At each position in a word there are at most two distinct squares whose last occurrence starts, and thus the following theorem holds (a short proof is given in [7]).

Theorem 2.1 [6]**.** *Any word of length n has at most $2n$ distinct squares.*

We now consider the one-hole case which behaves very differently from the zero-hole case. We will also count each square at the position where its last occurrence starts. If the last occurrence of a square in a partial word starts at position i, then it is a *square at position i*. In the case of partial words with one hole, there may be more than two squares that have their last occurrence starting at the same position. Such is the case with $a\diamond aababaab$ that has three squares at position 0: $(aa)^2$, $(aba)^2$ and $(abaab)^2$. Now, if we consider the word $a\diamond aababaabaaaa$, the square $(aa)^2$ has occurrences starting at positions 0 and 10. So $(aa)^2$ is not a square at position 0, but it is a square at position 10 since its last occurrence starts at position 10.

Theorem 2.2 [4]**.** *If a partial word with one hole has at least three distinct squares at the same position, then the hole is in the shortest square.*

The following lemma extends Fine and Wilf's periodicity result to partial words with one hole.

Lemma 2.1 [1]**.** *Let $w \in A_\diamond^*$ be weakly p-periodic and weakly q-periodic. If $H(w)$ is a singleton and $|w| \geq p + q$, then w is (strongly) $\gcd(p, q)$-periodic.*

The following lemmas on commutativity and conjugacy will also be useful for our purposes.

Lemma 2.2 [1]**.** *Let $x, y \in A^+$ and let $z \in A_\diamond^*$ be such that $H(z)$ is a singleton. If $z \subset xy$ and $z \subset yx$, then $xy = yx$.*

Lemma 2.3 [5]**.** *Let $x, y, z \in A_\diamond^*$ be such that $|x| = |y| > 0$. Then $xz \uparrow zy$ if and only if xzy is weakly $|x|$-periodic.*

Lemma 2.4 [3]**.** *Let $x, y \in A_\diamond^+$ and $z \in A^*$. If $xz \uparrow zy$, then there exist $v, w \in A^*$ and an integer $n \geq 0$ such that $x \subset vw$, $y \subset wv$, and $z = (vw)^n v$. Consequently, if $xz \uparrow zy$, then xzy is (strongly) $|x|$-periodic.*

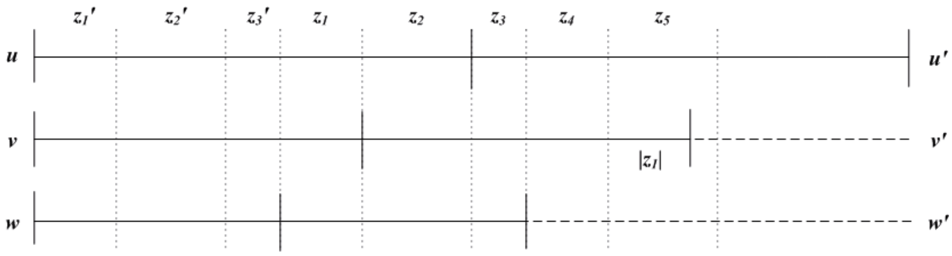

FIGURE 1. The case when $|ww'| > |u|$.

Theorem 2.3. *Let ww', vv' and uu' be three squares at the same position, with $|w| < |v| < |u|$. If $H(uu')$ is a singleton, then $|ww'| \leq |u|$.*

Proof. Since $|w| < |v| < |u|$, let us denote $v = wz_1$ and $u = vz_2$, for some partial words z_1, z_2 over the alphabet A. By contradiction, let us assume that $|ww'| > |u|$, and denote $ww' = uz_3$, where $z_3 \in A_\diamond^*$. According to Theorem 2.2, the hole is in ww'. We have $w' = z_1z_2z_3$, $w = z_1'z_2'z_3'$, $v = z_1'z_2'z_3'z_1$ and $u = z_1'z_2'z_3'z_1z_2$, where $z_i' \uparrow z_i$ for all $i \in \{1,2,3\}$. Since $v \uparrow v'$, we get that there exists $z_4 \in A^*$ such that $z_2z_3z_4$ is a prefix of v' and $|z_4| = |z_1|$, and by looking at the prefixes of length $|w|$ of u and u', we get that there exists $z_5 \in A^*$, with $|z_5| = |z_2|$, such that $z_1'z_2'z_3' \uparrow z_3z_4z_5$ (see Fig. 1). There are six cases to consider: Case 1 (respectively, Case 2, Case 3, Case 4, Case 5, Case 6) where the hole is in z_3 (respectively, z_2, z_1, z_3', z_2', z_1'). We treat Cases 1, 4 and 6 (Cases 2 and 3 are similar to Case 1, and Case 5 to Case 4).

Case 1. The hole is in z_3.

We have $w' = z_1z_2z_3$, $w = z_1z_2z_3'$, $v = z_1z_2z_3'z_1$ and $u = z_1z_2z_3'z_1z_2$ where $z_3 \subset z_3'$. Since $z_1z_2z_3', z_2z_3z_4$ are prefixes of v and v' respectively and $|z_1z_2z_3'| = |z_2z_3z_4|$, we get $z_1z_2z_3' \uparrow z_2z_3z_4$, and $z_1z_2z_3 \uparrow z_2z_3z_4$ by weakening. By Lemma 2.3, we get

$$z_1z_2z_3z_4 \text{ is weakly } |z_1|\text{-periodic.} \tag{2.1}$$

Now, since w is without holes, the prefixes of length $|w|$ of v and v', and respectively of u and u' are compatible, and $z_2z_3z_4 \subset w$ and $z_3z_4z_5 \subset w$, we get $z_2z_3z_4 \uparrow z_3z_4z_5$. Using Lemma 2.3 again, we get

$$z_2z_3z_4z_5 \text{ is weakly } |z_2|\text{-periodic.} \tag{2.2}$$

Finally, applying the weakening rule for the prefixes of length $|w|$ of u and u', we get $z_1z_2z_3 \uparrow z_3z_4z_5$. After using Lemma 2.3, we get

$$z_1z_2z_3z_4z_5 \text{ is weakly } |z_1z_2|\text{-periodic.} \tag{2.3}$$

From (2.1) and (2.3) we get that $z_1z_2z_3z_4$ is weakly $|z_1|$- and weakly $|z_1z_2|$-periodic. Applying Lemma 2.1, we get that $z_1z_2z_3z_4$ is $\gcd(|z_1|, |z_1z_2|)$-periodic. Hence there exists a word $x \in A^*$ of length $\gcd(|z_1|, |z_1 z_2|)$, such that $z_1 = x^m$ and $z_1 z_2 = x^{m+n}$ for some integers $m, n > 0$.

From (2.2) and (2.3) we get that $z_2 z_3 z_4 z_5$ is weakly $|z_2|$- and weakly $|z_1 z_2|$-periodic. Applying Lemma 2.1, we get that $z_2 z_3 z_4 z_5$ is $\gcd(|z_2|, |z_1 z_2|)$-periodic. Since $\gcd(|z_1|, |z_1 z_2|) = \gcd(|z_2|, |z_1 z_2|)$, we get that $z_2 z_3 z_4 z_5$ is $|x|$-periodic. Because $|z_1| \geq |x|$ and $|z_2| \geq |x|$ we get that $z_1 z_2 z_3 z_4 z_5$ is $|x|$-periodic.

Because z_1 and z_5 share a prefix of length $\min(|x^m|, |x^n|)$ with $m, n > 0$, z_5 is $|x|$-periodic and $|z_5| = |x^n|$, we get that $z_5 = x^n = z_2$. Since $z_3 z_4 z_5$ is $|x|$-periodic, $|z_5| \geq |x|$ and $|z_4| = |x^m|$, we get that $z_4 = x^m = z_1$.

Since $z_1 z_2 z_3 z_4$ is $|x|$-periodic and $z_1 z_2 = x^{m+n}$, it results that $z_3 \subset (x'x'')^p x'$ and $z_4 = (x''x')^m$ where $x = x'x''$ and $p \geq 0$ is an integer. But $z_4 = z_1 = x^m$. Hence, $x'x'' = x''x'$ and there exists a word y, such that $x' = y^q$ and $x'' = y^r$ for some integers $q, r \geq 0$.

Since $v' \uparrow v$, we have that $z_2 z_3 z_1 z_1 \uparrow z_1 z_2 z_3' z_1$. By cancellation, we get $z_2 z_3 z_1 \uparrow z_1 z_2 z_3'$. Replacing z_1 by x^m and z_2 by x^n, we get $x^n z_3 x^m \uparrow x^m x^n z_3'$, and consequently $z_3 x^m \uparrow x^m z_3'$ by cancellation. By Lemma 2.4, there exist words y', y'' such that $z_3 \subset y'y''$, $z_3' = y''y'$, and $x^m = (y'y'')^r y'$ for some integer $r \geq 0$. By Lemma 2.2, since $z_3 \subset y'y''$ and $z_3 \subset z_3' = y''y'$, we get $y'y'' = y''y'$. The latter implies that there exists a word z such that y' and y'' are powers of z. We obtain $x^m = z^{m'}$ for some integer m', and x and z are hence powers of a common word z'. We conclude that z_1, z_2, z_3, z_3', z_4 and z_5 are contained in powers of z', implying that there is a later occurrence of a square compatible with w^2.

Case 4. The hole is in z_3'.

Looking at the prefixes of length $|w|$ of v and v', we have $z_1 z_2 z_3' \uparrow z_2 z_3 z_4$. Applying weakening and Lemma 2.3, we get that $z_1 z_2 z_3' z_4$ is weakly $|z_1|$-periodic. Also, by looking at the prefixes of length $|w|$ of u and u' we get that $z_1 z_2 z_3' \uparrow z_3 z_4 z_5$. We apply weakening and Lemma 2.3 again, and get that $z_1 z_2 z_3' z_4 z_5$ is weakly $|z_1 z_2|$-periodic. Using Lemma 2.1, it follows that $z_1 z_2 z_3' z_4$ is $\gcd(|z_1|, |z_1 z_2|)$-periodic. Hence, there exists x such that $z_1 = x^m$ and $z_1 z_2 = x^{m+n}$, for some positive integers m, n, with $|x| = \gcd(|z_1|, |z_1 z_2|)$. There exist x', x'' such that $x = x'x''$, $z_3' \subset (x'x'')^p x'$, for some integer $p \geq 0$, and $z_4 = (x''x')^m$. Since the hole is in z_3', either there are integers p_1, p_2 and word x_1' having one hole such that $z_3' = (x'x'')^{p_1} x_1' (x''x')^{p_2}$ with $x_1' \subset x'$ and $p_1 + p_2 = p$, or there are integers p_1, p_2 and word x_2' having one hole such that $z_3' = (x'x'')^{p_1} x' x_2' (x'x'')^{p_2} x'$ with $x_2' \subset x''$ and $p_1 + p_2 + 1 = p$. Because $z_3' \subset z_3$, it implies that either $z_3 = (x'x'')^{p_1} x_1 (x''x')^{p_2}$ with $x_1' \subset x_1$, or $z_3 = (x'x'')^{p_1} x' x_2 (x'x'')^{p_2} x'$ with $x_2' \subset x_2$.

But also, $z_1 z_2 z_3' \uparrow z_2 z_3 z_4$. Hence, we get that $x^m z_3' \uparrow z_3 z_4$. This is equivalent to one of the following cases:

$$x^m (x'x'')^{p_1} x_1' (x''x')^{p_2} \uparrow (x'x'')^{p_1} x_1 (x''x')^{p_2} (x''x')^m$$

when we get $x_1 = x'$, or

$$x^m (x'x'')^{p_1} x' x_2' (x'x'')^{p_2} x' \uparrow (x'x'')^{p_1} x' x_2 (x'x'')^{p_2} x' (x''x')^m$$

when we get $x_2 = x''$. In either case, $z_3 = (x'x'')^p x'$.

Since $z_1 z_2 z_3' \uparrow z_3 z_4 z_5$, there is the possibility that $z_5 = (x''x')^n$ if $n \leq p_2$. We leave this case to the reader and assume that $n > p_2$. We get that $z_5 = (x''x')^{n_1} x'' x_1 (x''x')^{n_2}$ with $x_1' \subset x_1$, or $z_5 = (x''x')^{n_1} x_2 x' (x''x')^{n_2}$ with $x_2' \subset x_2$ (in either case $n_1 + n_2 + 1 = n$). Since $v \uparrow v'$, it follows that z_5 and z_1 share a prefix of length $|x|$, and so z_5 has $x'x''$ as a prefix. There are three cases to consider: (4.1) $x'x'' = x''x'$; (4.2) $x'x'' = x''x_1$; and (4.3) $x'x'' = x_2 x'$. For (4.1), there exists a word y such that x' and x'' are powers of y. It follows that z_1, z_2, z_3, z_3' and z_4 are contained in powers of y, implying that there is a later occurrence of a partial word that is compatible with the square $(w')^2$. For (4.2) and (4.3), $n_1 = 0$ and we can denote z_5 as $x'x''(x''x')^{n-1}$. Furthermore, since $z_1 z_2 z_3' \uparrow z_3 z_4 z_5$ we get that either $x^{m+n+p_1} x_1' (x''x')^{p_2} \uparrow x^{m+p} x' x' x'' (x''x')^{n-1}$ or $x^{m+n+p_1} x' x_2' (x'x'')^{p_2} x' \uparrow x^{m+p} x' x' x'' (x''x')^{n-1}$. We prove the first case (the other is similar).

If $p > n + p_1$, then $p_2 > n$ a contradiction. If $p = n + p_1$, then $x_1' x'' x' \uparrow x' x' x''$ and $x'x'' = x''x'$, the same contradiction as before follows. If $p < n + p_1$, we get $x^{n-p_2} x_1' \uparrow x' x' x'' (x''x')^{n-1-p_2}$. If $p_2 < n - 1$, then again $x'x'' = x''x'$. If $p_2 = n - 1$, then $x'x'' \uparrow x'' x_1'$. By Lemma 2.4, there exist words y', y'' such that $x' = y'y''$, $x_1' \subset y''y'$, and $x'' = (y'y'')^r y'$ for some integer $r \geq 0$. By Lemma 2.2, since $x_1' \subset y''y'$ and $x_1' \subset x' = y'y''$, we get $y'y'' = y''y'$. The latter implies that there exists a word z such that y' and y'' are powers of z. We obtain x' and x'' are powers of z. We conclude that z_1, z_2, z_3, z_3', z_4 and z_5 are contained in powers of z, implying that there is a later occurrence of a square compatible with $(w')^2$.

Case 6. The hole is in z_1'.

Since $v \uparrow v'$, $z_1' z_2 z_3$ and $z_2 z_3 z_4$ are prefixes of v and v' respectively, and $|z_1' z_2 z_3| = |z_2 z_3 z_4|$, we have that $z_1' z_2 z_3 \uparrow z_2 z_3 z_4$. Applying Lemma 2.4, we get that $z_1' z_2 z_3 z_4$ is $|z_1|$-periodic. Since $u \uparrow u'$, we get that $z_1' z_2 z_3 \uparrow z_3 z_4 z_5$, and $z_1' z_2 z_3 z_4 z_5$ is $|z_1 z_2|$-periodic. Hence, $z_1' z_2 z_3 z_4$ is $|x| = \gcd(|z_1|, |z_1 z_2|)$-periodic, where $z_1' \subset x^m$ and $z_2 = x^n$, for some word x and integers $m, n > 0$. This implies that there exist $x', x'' \in A^*$, such that $x = x'x''$, and $z_3 = (x'x'')^p x'$, for some integer $p \geq 0$, $z_4 = (x''x')^m$ and $z_5 = (x''x')^n$ (because $z_2 z_3 z_4 z_5$ is $|z_1 z_2|$-periodic and $|z_2 z_3 z_4| > |z_1 z_2|$).

Since $v \uparrow v'$, if $|z_2| \geq |z_1|$, then z_5 and z_1 share a common prefix of length $|x^m|$. It follows that $z_1 = (x''x')^m$. But, since $z_1' \subset z_1$, it results that $z_1' \subset (x''x')^m$, and recall that $z_1' \subset (x'x'')^m$. If $m > 1$, then we get $x'x'' = x''x'$ and so there exists y, such that $x' = y^q$ and $x'' = y^r$ for some nonnegative integers q, r, giving us a contradiction with the assumption that there is no later occurrence of a factor compatible with $(w')^2$. If $m = 1$, then we also get $x'x'' = x''x'$ by Lemma 2.2. Hence, we may assume that $|z_2| < |z_1|$ and z_1 has as a prefix $(x''x')^n$. Let $z_6 \in A^*$, where $|z_6| = |z_2|$, such that $z_1' z_2 z_3 z_1 \uparrow z_3 z_4 z_1 z_6$ (the prefixes of length $|v|$ of u and u' are compatible). By using simplification we get that $z_2 x' z_1 \uparrow x' z_1 z_6$ and $z_2 x' z_1 z_6$ is $|z_2|$-periodic by Lemma 2.4. Since $|z_1| = |x^m| > |z_2|$, it follows that $z_1 = (x''x')^m$. Since $z_1' \subset (x'x'')^m$, we get a contradiction as before.

Since Cases 1–6 lead to contradiction we conclude that $z_3 = \varepsilon$. □

Let us now assume that the hole is at a position i in a word of length n. The upper bound for the maximum number of factors, compatible with distinct squares, would be achieved if all these factors, starting before the hole, would contain the hole (then more than two squares can start at the same position). Note that in the case when a square containing a hole has its last occurrence at a certain position, no other word that is a square can have its last occurrence starting at the same position (otherwise a later occurrence of the same word, or a word compatible with it, would appear later in the word). Let us look at the start position j of a square containing the hole and denote it as j (obviously, there are at most i such squares). Let us denote the length of such square by n_j.

Hence, if at position j we have a square of length n_j, then according to Theorem 2.3, up to position $2n_j + j$ we will have counted at most three distinct squares. Using an induction we notice that up to position $2^m n_j + j$ we will have counted at most $2m + 1$ distinct squares. Since the length of the word is n, we have that the maximum value for m, for squares starting at position j, is bounded by $\log(\frac{n-j}{n_j})$.

Note that the maximum is achieved for the case when n_j is minimum. Hence we can replace in our formula n_j by $i - j$, which is the smallest length a square starting at position j and containing the hole may have.

Theorem 2.4. *The number of distinct squares compatible with factors in any partial word with one hole of length n is at most* $\frac{7n}{2}$.

Proof. Using the previous remarks it is easy to see that the number of squares at position j and containing the hole is $2\log(\frac{n-j}{i-j}) + 1$. Hence, we get that the total number of distinct squares that we can obtain is

$$\sum_{j=0}^{i-1}(2\log(\tfrac{n-j}{i-j}) + 1) = i + \tfrac{2}{\ln 2}\sum_{j=0}^{i-1}\ln(\tfrac{n-j}{i-j}).$$

The sum from the previous formula is equal to $\sum_{x=n-i+1}^{n}\ln(x) - \sum_{y=1}^{i}\ln(y)$ and implicitly, less or equal than $\int_{n-i+1}^{n+1}\ln(x)\mathrm{d}x - \int_1^i \ln(y)\mathrm{d}y$. After integrating we get

$$(n+1)\ln(n+1) - (n-i+1)\ln(n-i+1) - i\ln(i).$$

Since the maximum is obtained for $i = \frac{n+1}{2}$, the function is hence less than $(n+1)\ln 2$. Using Theorem 2.1 for the rest of the word, we get that the number of distinct squares, compatible with factors of the word, is bounded by

$$2n - \tfrac{n+1}{2} + \tfrac{2}{\ln 2}(n+1)\ln 2 = \tfrac{7n}{2} + \tfrac{3}{2}\cdot$$

Since the last position in the word contains no squares, we get that the maximum number of factors compatible with distinct squares is smaller than $\frac{7n}{2}$. □

This bound can be slightly improved by using Ilie's $2n - \Theta(\log n)$.

3. Conclusion

In order to improve the bound stated in Theorem 2.4, we need to somehow limit to less than 3.5 the average number of squares that have their last occurrence starting at the positions of the partial word. This requirement draws attention to positions i where three or more squares have their last occurrences. Is it true that at positions "neighbouring" to i, no squares can have their last occurrences? In fact, if at position i we have at least three factors compatible with squares, this does not imply that at position $i+1$ we will have less. Indeed, consider the example

$$ab\diamond abcabbeabcabdabcabbeabcabbabcabbeabcabdabcabbeabcab$$

where at position 0 we have factors compatible with the squares

$$(abc)^2,\ (abdabcabbeabc)^2 \text{ and } (abbabcabbeabcabdabcabbeabc)^2$$

at position 1 factors compatible with the squares

$$(bca)^2,\ (beabcab)^2,\ (bdabcabbeabca)^2 \text{ and } (bbabcabbeabcabdabcabbeabca)^2$$

and at position 2 factors compatible with

$$a^2,\ (cab)^2,\ (dabcabbeabcab)^2 \text{ and } (babcabbeabcabdabcabbeabcab)^2.$$

In [8], Ilie gave a relation between the lengths of squares at positions neighbouring a position where two squares have their last occurrences. More precisely, he showed that if $v^2 < u^2$ are two squares at position i and w^2 is a square at position $i+1$, then either $|w| \in \{|v|, |u|\}$ or $|w| \geq 2|v|$ (see Lem. 2 of [8]). Referring to the above example, we observe that such is not the case with partial words with one hole.

References

[1] J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.* **218** (1999) 135–141.

[2] F. Blanchet-Sadri, *Algorithmic Combinatorics on Partial Words*. Chapman & Hall/CRC Press, Boca Raton, FL (2008).

[3] F. Blanchet-Sadri and D.K. Luhmann, Conjugacy on partial words. *Theoret. Comput. Sci.* **289** (2002) 297–312.

[4] F. Blanchet-Sadri, R. Mercaş and G. Scott, Counting distinct squares in partial words, edited by E. Csuhaj-Varju, Z. Esik, *AFL 2008, 12th International Conference on Automata and Formal Languages, Balatonfüred*, Hungary (2008) 122–133, www.uncg.edu/cmp/research/freeness

[5] F. Blanchet-Sadri, D.D. Blair and R.V. Lewis, Equations on partial words. *RAIRO-Theor. Inf. Appl.* **43** (2009) 23–39, www.uncg.edu/cmp/research/equations

[6] A.S. Fraenkel and J. Simpson, How many squares can a string contain? *J. Combin. Theory Ser. A* **82** (1998) 112–120.

[7] L. Ilie, A simple proof that a word of length n has at most $2n$ distinct squares. *J. Combin. Theory Ser. A* **112** (2005) 163–164.

[8] L. Ilie, A note on the number of squares in a word. *Theoret. Comput. Sci.* **380** (2007) 373–376.

Communicated by C. Choffrut.
Received July 23, 2008. Accepted August 18, 2009.